Protein Science (1999), *8*:1181–1190. Cambridge University Press. Printed in the USA.

REVIEW

Folding funnels, binding funnels, and protein function

CHUNG-JUNG TSAI,[1] SANDEEP KUMAR,[2] BUYONG MA,[1] AND RUTH NUSSINOV[2,3]

[1]Laboratory of Experimental and Computational Biology, NCI-FCRDC, Bldg. 469, Room 151, Frederick, Maryland 21702
[2]Intramural Research Support Program—SAIC, Laboratory of Experimental and Computational Biology, NCI-FCRDC, Bldg. 469, Room 151, Frederick, Maryland 21702
[3]Sackler Institute of Molecular Medicine, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

(RECEIVED January 22, 1999; ACCEPTED March 4, 1999)

Abstract

Folding funnels have been the focus of considerable attention during the last few years. These have mostly been discussed in the general context of the theory of protein folding. Here we extend the utility of the concept of folding funnels, relating them to biological mechanisms and function. In particular, here we describe the shape of the funnels in light of protein synthesis and folding; flexibility, conformational diversity, and binding mechanisms; and the associated binding funnels, illustrating the multiple routes and the range of complexed conformers. Specifically, the walls of the folding funnels, their crevices, and bumps are related to the complexity of protein folding, and hence to sequential vs. nonsequential folding. Whereas the former is more frequently observed in eukaryotic proteins, where the rate of protein synthesis is slower, the latter is more frequent in prokaryotes, with faster translation rates. The bottoms of the funnels reflect the extent of the flexibility of the proteins. Rugged floors imply a range of conformational isomers, which may be close on the energy landscape. Rather than undergoing an *induced fit* binding mechanism, the conformational ensembles around the rugged bottoms argue that the conformers, which are most complementary to the ligand, will bind to it with the equilibrium shifting in their favor. Furthermore, depending on the extent of the ruggedness, or of the smoothness with only a few minima, we may infer nonspecific, broad range vs. specific binding. In particular, folding and binding are similar processes, with similar underlying principles. Hence, the shape of the folding funnel of the monomer enables making reasonable guesses regarding the shape of the corresponding binding funnel. Proteins having a broad range of binding, such as proteolytic enzymes or relatively nonspecific endonucleases, may be expected to have not only rugged floors in their folding funnels, but their binding funnels will also behave similarly, with a range of *complexed* conformations. Hence, knowledge of the shape of the folding funnels is biologically very useful. The converse also holds: If kinetic and thermodynamic data are available, hints regarding the role of the protein and its binding selectivity may be obtained. Thus, the utility of the concept of the funnel carries over to the origin of the protein and to its function.

Keywords: binding funnels; conformational ensembles; energy landscape; folding funnels; function; misfolding

The concept of folding funnels, which has been conceived a number of years ago (Bryngelson & Wolynes, 1989; Karplus & Shakhnovitch, 1992; Baldwin, 1994, 1995; Karplus et al., 1995; Onuchic et al., 1995; Wolynes et al., 1995; Dill & Chan, 1997; Karplus, 1997; Lazaridis & Karplus, 1997; Gruebele & Wolynes, 1998), has revolutionized our understanding of protein folding. Its most important point, namely, the stipulation that protein folding progresses via multiple routes going downhill rather than through a single pathway, has immediately elegantly shown a way out of the long-standing baffling Levinthal paradox (Levinthal, 1969). Furthermore, the funnel concept has illustrated how the down-gliding conformation may get trapped at some crevice along its way and, depending on the depth of its trap and the height of its barrier-bump, its gliding would be resumed. More recent work has shown that some modification of the multiple pathways descending downhill may be in place (Gruebele & Wolynes, 1998; Martinez et al., 1998). While there are many conformations going through numerous paths, there still appear to be some obligatory steps in the folding reaction. Transition state ensembles may be well defined and conformationally restricted (Martinez et al., 1998). The recent illustration that the different funnel energy landscapes can be correspondingly portrayed by disconnectivity graphs pictorially drawn as different types of rooted trees (Frauenfelder et al., 1991; Becker & Karplus, 1997; Frauenfelder & Leeson, 1998; Wales et al., 1998) has been very instrumental. These trees aid in understanding how, despite the significant traps that the down-marching conformations encounter, they still manage to reach the bottom of the funnel. Yet,

Reprint requests to: R. Nussinov, NCI-FCRF Bldg. 469, Room 151, Frederick, Maryland 21702; e-mail: ruthn@ncifcrf.gov.

in spite of the inspiring beauty of the folding funnel concept, its implications for protein function have not been considered. Proteins, like any other molecule in vivo or in vitro, function through their binding. Hence, to understand protein function, we need to consider their intra- and intermolecular associations.

Recently we have described the implications of folding funnels to straightforwardly rationalize binding mechanisms. Considerable attention has been focused on the slopes of the folding funnels, their bumps and crevices, and their bottoms. Yet, the protein molecule always functions through binding. And, when considering binding mechanisms, the unbound protein has been largely viewed as existing in a single (most stable) conformation, hence, the terms "lock-and-key" and "induced-fit," "crystal effects," etc. On the other hand, when viewed in the context of conformational isomers around the bottom, binding mechanisms such as crystallization, induced-fit, specific vs. broad-range nonspecific binding, domain swapping (Bennett et al., 1994, 1995), and misfolding are explained simply by the extent of the ruggedness of the funnel around the bottom, the narrow valleys, and the barrier heights. A rigid protein, with a highly specific binding, can be viewed as having a smooth bottom with a single or very few minima. On the other hand, a nonspecific protein, binding to a range of potential ligands can be pictured as one having a very rugged bottom with rather low barriers separating the multiple minima valleys. We have therefore argued that the ensemble of conformational isomers around the bottom of the folding funnel implicitly replace long-held notions in binding, such as "lock-and-key" (Fischer, 1894) and "induced-fit" (Koshland, 1958), crystal packing effects, hinge bending motions, domain swapping (Bennett et al., 1994, 1995), and misfolding. Depending on the ruggedness of the folding funnel around its bottom, its hills and canyons, and their corresponding heights and depths, these can be directly understood. The more flexible the protein, the larger the ensemble of conformers, and the lower the barriers between them. The conformer that binds the ligand is the one that is complementary to it, with the equilibrium adjusting itself in favor of this conformer (Foote & Milstein, 1994).

Previously, we have presented a general scheme for protein folding and binding (Tsai et al., 1996, 1997a, 1997b; Xu et al., 1997, 1998). That scheme has evolved from a series of systematic investigations encompassing protein architecture, the hydrophobic effect, hydrophilic bridges, compact hydrophobic folding units, and domain swapping (Xu et al., 1998). The clear and uniform conclusion was that the sole difference between folding and binding is the presence, or absence, of chain connectivity. Hence, by recognizing the similarities, and the differences, between these two processes, they can be utilized toward understanding, and hence toward prediction of both.

Discussion

Macrostate, microstate, and microfunnel-like energy landscape

The new view of protein folding, embodied in the funnel-shaped energy landscape, is a beautiful conceptual base capturing the realities of the folding process. Its attractiveness lies in its immediate utility: Already the bare concept of the funnel shape explains why protein folding is neither a random search nor does it follow a single pathway toward its native conformation. In this back-to-back review series, the profound first review (Dill, 1999) adds a new statistical mechanics description to the funnel-shaped energy landscape theory. In the second review, we provide a practical folding/binding model within the framework of the funnel concept, describing its usefulness for the mechanism of protein folding, binding, and function.

The language used in the funnel-shaped energy landscape is implicitly referred to as a macrostate, that is, an ensemble as described in statistical mechanics (Dill, 1999). A single pathway racing down to the bottom of a funnel, which can only be observed in a computer simulation, is a micropath. Now, in the preceding review, Dill has filled the gap between the macrostate and the microstate. Bridging between the microscopics and macroscopics further validates the notion and the proposition that the energy landscape concept can be successfully used for describing folding thermodynamics and kinetics (Chan & Dill, 1994, 1998), as well as for developing faster search algorithms (Dill et al., 1997).

In this review, we start with a practical description of the micropath, a single pathway rolling down toward the funnel bottom. In our model (Tsai et al., 1998, 1999), protein folding is the outcome of a combinatorial assembly of a set of transient building blocks. The formation of any building block in a given sequence can also already be described and guided by a microfunnel-like energy landsacpe. The mutual recognition between building blocks resembles a fusion of two microfunnel-like landscapes. At the bottom of a subfunnel-like landscape resides a compact, stable hydrophobic folding unit. Such a hydrophobic folding unit, in turn, serves as the basic unit in building a functional, multidomain protein, an oligomer, or a functional complex.

Here we discuss and relate distinctive funnel shapes and surface ruggedness to a variety of protein folding, binding, and functional mechanisms. We focus on the shapes of the energy landscape associated with protein complexes and their implications. In binding, as in folding, multiple conformations may race down the funnel toward favorable, complexed structures. Nevertheless, due to the fact that the number of atoms is larger in complexes as compared to protein monomers, and due to the lack of the chain connectivity, the potential number of conformations in binding is substantially larger than in folding. Binding funnels can therefore be very complicated.

Protein binding

Proteins, like any other molecule in vivo or in vitro, function through their binding. The target molecules vary in size and complexity from a single metal cation to proteins or other macromolecules. Protein binding is comparable to protein folding only when the ligands are amino acids, peptides, or proteins. However, there is no difference in the principles governing binding of different types of molecules, such as amino acids or nucleotides. For this reason, we do not specify the chemical nature of the ligands explicitly, and our discussion is general in nature.

Protein binding may be classified utilizing different criteria. These may include the chemical nature of the ligands. Alternatively, it may be the size of the ligand, i.e., a small molecule or macromolecular binding. Further, considering the outcome of the interaction process, binding may be classified as "inert" or "reactive." Inert binding implies simple physical binding with no chemical modification of either the protein or the ligand. On the other hand, reactive binding implies that a chemical reaction takes place as a result of the binding process, as in the case of an enzyme catalyzed reaction. Binding processes may further be classified by the stability of the protein as a monomer. In a two-state binding

process, the protein has a single native state, which is the bound state. In a three-state binding process, the protein is stable in both monomeric and bound states. Binding may further be classified by the rigidity of the protein and of the ligand. Carrying out a careful examination of the energy landscapes of different binding processes and comparing them with folding funnels may therefore prove beneficial. For example, as we will see later, rigidity and flexibility of either the free or the bound protein are closely related to protein function.

Folding funnels and binding funnels: Concept

Protein folding is a spontaneous process. It is executed in the cell at a constant temperature, below its transition temperature. As protein folding behaves like a first-order two-state phase transition, both states (the native conformation and the associated, unfolded state) have equal population times at the transition temperature. Below the transition temperature, the native folded protein is thermodynamically more favorable than the unfolded state, which is an ensemble of denatured conformations. By utilizing a funnel shape to portray the protein folding energy landscape as a function of conformational space, one implies that protein folding is a self-driven process rather than a random search. Along the surface of a funnel, there are many alternate paths flowing from the top toward its bottom. Folding kinetics depends on how long an individual descending conformation will be trapped at some bumps on the funnel surface. In a two-state transition, free energy barriers, which are the source of the bumps, are expected along the funnel surface. Clearly, the higher the barrier, the longer it will take the downhill folding process to reach the native conformation near the bottom of the funnel. The funnel theory provides particularly attractive guidelines for describing the protein folding process. Nevertheless, a "practical" folding model is still essential both to explain the variability in the behavior of protein folding, with some molecules folding rapidly, and others slowly, and some more prone to misfolding than others, and to clarify the logical consequences for binding.

The process of protein folding can be either sequential or nonsequential (Tsai et al., 1999). In a sequentially folding protein, sequentially covalently connected fragments are adjacent to one another in the three-dimensional (3D) structure. Such folding is more often seen in eukaryotes and is thought to be the outcome of slower translation rates (Netzer & Hartl, 1997). In contrast, in a nonsequentially folded protein, structurally adjacent fragments are not connected sequentially. As a result, in such proteins the polypeptide chain crosses back and forth between different domains (or hydrophobic folding units). Nonsequential folding is more common in prokaryotes typically having faster translation rates. The landscape of folding funnels can be smooth and simple. Simpler funnels are likely to be observed in sequentially folding proteins. Nonsequentially folding proteins may translate to complicated funnels, with crevices and bumps along their paths. Such proteins have a higher probability of being trapped in local minima, and hence misfold. On the other hand, sequentially folding proteins, in which consecutive fragments in the chain interact with each other are less likely to misfold. Figure 1 depicts this relationship between the complexity of protein folding and the shape of the funnels.

During evolution, proteins with nonrandom sequences, which are able to fold rapidly, within biologically relevant time scales, have been selected. Folding funnels for such proteins have funnel-like energy landscapes. These landscapes consist of deep valley(s) with well-defined local and global minima. On the other hand, a protein with a random sequence will not be able to fold rapidly, due to the lack of stabilizing intramolecular hydrophobic and electrostatic interactions. Since for such a random sequence protein there is no energy difference between ordered and disordered states, this protein is expected to have a very shallow funnel (*if* it exists at all), lacking any well-defined local or global minimum. As such a protein is unlikely to exist in the living cell, we may argue that folding funnels are the result of evolutionary selection, acting in accord with physical laws.

Clearly, if we use the unfolded state as a reference state and all degrees of freedom as coordinates, a binding process involving an already folded protein can be depicted by a funnel-like shape. Further, there is also no doubt that the less frequent case of a two-state binding process, in which binding and folding take place simultaneously, also displays a funnel-like shape. However, as it is more convenient to use the folded protein as a reference state, the question arises as to whether the binding process can still be portrayed by a funnel under these circumstances. In the extreme case of rigid protein binding, the degrees of freedom are reduced to three translational and three rotational for each two-molecule association. However, even rigid docking requires the robust determination of the global free energy minimum from the myriad of all possible conformations. In this sense, problems of protein folding and binding are quite similar, and thus one can extend Levinthal's paradox to protein binding. However, as in the case of folding, it can also be readily explained by the existence of the multiple pathways, here for protein–protein binding. Hence, the extension of the funnel concept to binding processes should also be very useful (Miller & Dill, 1997).

The difficulties involved in the prediction of docked configurations have been a subject of numerous studies (Katchalski-Katzir et al., 1992; Cherfils & Janin, 1993; Norel et al., 1994, 1995; Fischer et al., 1995; Dunbrack et al., 1997). The problem can be defined as follows: given the structures of two molecules, predict their "correct" bound configuration. If the structures to be docked are taken from their crystallized cocomplex, predicting their fitted association is relatively straightforward. On the other hand, if the structures have been determined when the molecules are in their "free," "unbound" state, the problem can be very difficult. Yet, even structures presumably determined in the unbound state are actually also bound; except in this case they are bound to their twin molecule in the crystal. If the molecules are relatively rigid and have relatively smooth funnels with a single or a few minima, their conformational diversity is limited, and hence there is a higher likelihood that the conformers binding in the "free," "unbound" state will be those binding to their ligand(s). If, however, the protein is flexible, its folding funnel has rugged bottom consisting of several minima separated by small energy barriers. Hence, its conformational diversity is larger, and the conformer binding to its cofactor, or ligand, may be different from the one which is bound in the crystal. In such a case, if in predicting the conformation of the complex one picks the structure that is found in the crystal of the single molecule, achieving a correct prediction of the bound conformation is much more difficult. As Verkhivker et al. (1996) have pointed out, for robust ligand-protein docking, the binding energy landscape must have a funnel shape, leading to the global free energy minimum. This has recently been elegantly illustrated by Zhang et al. (1999), who have shown that there are energy gradients, or funnels, near the binding sites. They have further used their finding to provide an explanation for the rapid association rates.

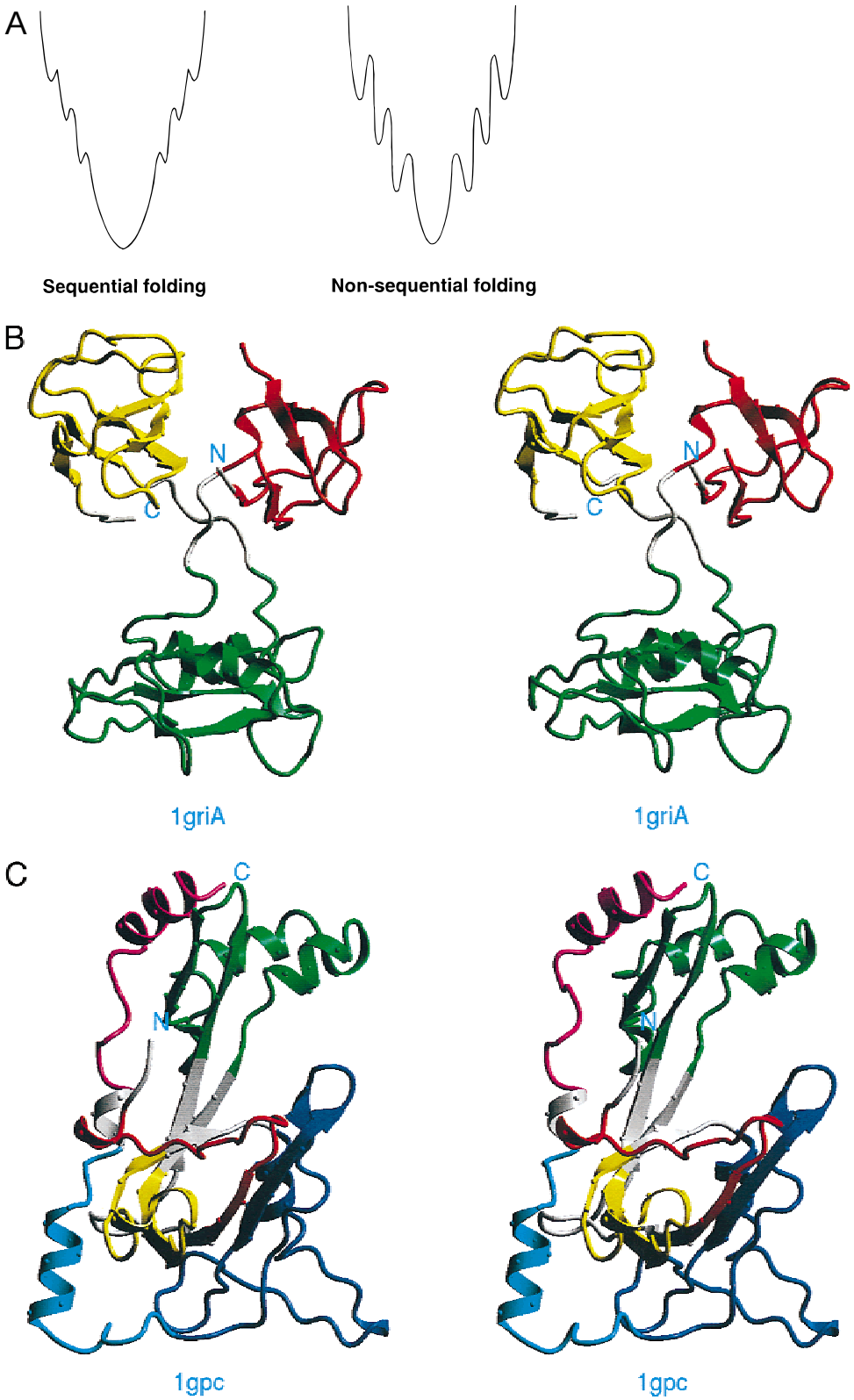

Fig. 1. The figure depicts the relationship between the complexity of protein folding and the shape of the funnels. **A:** The concept of folding funnels in terms of sequential folding vs. nonsequential folding. Based on merely the entropy consideration, proteins that fold sequentially are expected to have a smoother, less bumpy surface than nonsequentially folded proteins. **B,C:** Examples of proteins corresponding classified as a sequentially folded protein (Protein Data Bank code 1GRI, growth factor bound protein 2) and a nonsequentially folded protein (1GPC, DNA-binding protein core GP32). The folding complexity reflects the number of assigned "building blocks" (Tsai et al., 1998) and the order of their interactions. The sequential order of the ribbon colors starting from the N-terminus to the C-terminus are red, green, yellow, blue, magenta, and cyan. The unassigned segments are colored white.

As in folding, the multiple routes going downhill in binding rationalize how favorable protein associations may be reached in short times. In folding, the existence of multiple pathways gliding down the energy funnel implies that a single given protein molecule need not search through the entire conformational space. Instead, different protein molecules follow different pathways to reach the native state defined by the global minimum (minima). Hence, there is an ensemble of non-native conformations from which the folding process initiates and proceeds. Similar arguments can be advanced in rigid binding, with the multiple conformations going downhill circumventing the need for an exhaustive six-translational and rotational degrees of freedom search by a single binding conformation. Favorable binding may be achieved via multiple diffusion-collision processes of the two molecules (or, in folding, of the substructural component pairs). Hence, local optimization of the intermolecular interactions is likely to take place, with side-chain rotations and limited backbone movements. Simultaneously, alternate, parallel pathways will be manifested by the binding process. Owing to the absence of the constraint of chain connectivity in binding, more favorable associations may be obtained as compared to folding (Tsai et al., 1998).

The above arguments may be illustrated by the fact that protein–protein association generally occurs at rates which are 10^3 to 10^4 times faster than would be expected from simple considerations of collision frequencies ($\sim 10^9\ s^{-1}\ M^{-1}$; when considering orientation effects, $\sim 10^3\ s^{-1}\ M^{-1}$) (Northrup & Erickson, 1992; Wells, 1996). When proteins collide, they do not diffuse away immediately. Instead, they roll on one another and thereby sample considerably more surface area than would be the case for a single elastic collision. This is equivalent to multiple pathways and supports a funnel explanation.

Relationship between folding funnels and binding funnels in terms of rigidity and protein function

The funnels of rigid proteins are likely to display a single or a few minima. Such proteins may be highly specific in their binding. On the other hand, flexible proteins are likely to exist in solution in a range of conformations, yielding rugged funnel floors. For rigid proteins, geometrically similar structures are near each other on the energy landscape (Dill & Chan, 1997). If the individual funnels of the constituent protein monomers whose complexed association is explored are rugged, the funnel of the complex can be expected to be rugged as well. Thus, if the individual molecules exist in a range of conformational isomers, they will associate creating a range of complexed conformations. In principle, each of the single molecule conformers can associate, so long as favorable binding is obtained. On the other hand, the inverse does not necessarily hold. Rigid molecules displaying smooth funnels with a single minimum or a few minima may still associate in diverse ways, although probably to a lesser extent than the highly flexible ones. Part of this problem may be gleaned by analysis of electrostatic interactions in the proteins. Analyzing 294 salt bridges from a nonredundant dataset of 38 high resolution ($\leq$1.6 Å) crystal structures of dissimilar monomeric proteins, we have recently found that the majority (greater than three-quarters) of the salt bridges are formed within the hydrophobic folding units (Tsai & Nussinov, 1997). Only a small minority (less than one-tenth) of the salt bridges are formed across domains (or hydrophobic folding units). This study indicates that interdomain boundaries are frequently flexible and can allow hinge based motions. Our previous studies on hexameric glutamate dehydrogenase from hyperthermophilic and mesophilic organisms have shown similar trends. Moreover, salt bridges are much more frequent within the enzyme subunits than in intersubunit interfaces. Taken together, these results indicate that relatively rigid domains, or subunits, may yield binding funnels with rugged bottoms.

This straightforwardly explains the difficulty in the prediction of bound conformations if the monomer is flexible. In binding, the multiple pathways originate from the population of the already folded conformational isomers around the bottom of the respective folding funnels. Hence, the more flexible the monomers, the larger the number of conformers and the larger the number of alternate starting points of downhill pathways in the binding funnels. Depending on how favorable is the native conformation of the complex, and its rigidity, the multiple routes may yield an ensemble of conformers around the bottom of the binding funnel. The less stable molecular associations will have rugged binding funnel bottoms, with low barriers between the conformers of the complex.

While some proteins are highly selective, others may bind a broad spectrum of ligands. Proteins displaying specificity in their binding, recognize well-defined sequence structure motifs. Highly specific proteins are likely to be relatively rigid. In contrast, proteins binding to a range of ligands, are likely to be more flexible, with a population of diverse conformers. While one conformation fits one ligand, an alternate conformer may be more favorable for binding a ligand with a different structure (Foote & Milstein, 1994). The extent of the flexibility is likely to be related to the function of the protein. A specific antibody is likely to be more rigid than the less specific germline (Wedemayer et al., 1997). Similarly, some families of proteolytic enzymes, such as the aspartic proteinases, or the serine proteinases may be expected to display larger flexibility. Less specific enzymes, which cleave the DNA such as the topoisomerases, may similarly be expected to be quite flexible, as compared to the specific restriction endonucleases. Molecular flexibility enables the protein to bind to a range of potential ligands.

Flexible proteins have rugged funnel bottoms, corresponding to their range of conformational isomers, with low energy barriers separating them. Consistently, if the protein monomer is inherently flexible, its protein–protein complex is also likely to display a range of conformations, with relatively small differences in energy separating them. Hence, the bottom of the binding funnels of proteins whose functions dictate flexibility, is likely to be rugged, populated by ensembles of conformational isomers of *complexes*. This suggests that the bottom of the funnels of complexes of the less specific proteolytic enzymes would be relatively more rugged than those of the rigid proteins. Thus, knowledge of the protein function may be expected to provide hints about the shape of the bottom of its folding funnel, and correspondingly, about the shape of the bottom of its complexed, binding funnel. The converse also holds: Availability of kinetic data regarding the range of the conformational isomers of a protein, free or bound, is likely to provide hints about its type of function.

A recent example illustrates this point in a rather striking way. Lee et al. (1998) have determined the crystal structure of a recombinant plasmepsin II. This enzyme belongs to the family of aspartic proteases. There are two independent molecules in the asymmetric unit, displaying markedly different domain displacements. Furthermore, both molecules were complexed with pepstatin A, a general aspartic proteinase inhibitor, and a potent inhibitor of plasmepsin II. Here two pre-existing, different conformations of plasmepsin II bind pepstatin A, rather than the same, "initial" unbound confor-

mation binding through an induced fit mechanism. Hence, in a similar vein, the bottom of the funnel of the complex is populated by different complexed conformations. In the crystal, they bind to each other as different conformers. We may reasonably assume that additional conformational isomers of the complex exist in solution. However, the conformers of the complex that cocrystallize are more complementary, and hence more favorable for binding to each other. The equilibrium of the *complexed* conformations around the bottom of the binding funnel results in a conformational shift favoring the conformers of the complex, which bind to the growing crystal.

Previously, it has been shown that two molecules in the same crystallographic asymmetry unit can differ in their relative domain–domain orientation. Muller et al. (1998) have crystallized the tissue factor, a member of the cytokine receptor superfamily. Tissue factor is an obligate cofactor of coagulation factor VIIa. Muller et al. have shown that there is a hinge rotation of 12.7° in the rabbit tissue factor between the domains when the two molecules in the asymmetric unit are compared. This suggests that under crystallization conditions at least one of these conformers was in a highly populated state. The other conformer is complementary to it, and thus binds and cocrystallizes with it, with the equilibrium shifting in its direction. Hence, again, binding and folding behave similarly. In the plasmepsin case, the different conformers of the complex bind as a unit to the growing crystal. In the tissue factor case, it is different conformers of the *single molecules*, which bind to each other in the asymmetric unit of the crystal.

The function of the proteolytic enzymes is to catalyze hydrolysis of the protein backbone via nucleophilic attack on the carbonyl carbon of the peptide bond. During this nucleophilic attack, the trigonal carbonyl carbon of the peptide bond becomes tetragonal in the intermediate, or transition state (Creighton, 1993). Five different catalytic genera of proteolytic enzymes, in which serine, threonine, cysteine, aspartic or metallo groups provide a nucleophile during catalysis, are known. These five genera can be subdivided into innumerable clans and families (Barrett et al., 1998). For example, there are about 40 families of serine- and threonine-proteinases alone (Perona & Craik, 1995; Barrett et al., 1998). While being excellent examples of convergent evolution, these enzymes also show extensive structural diversity. Actually, it is often the case that enzymes belonging to different clans adopt different overall folds. Consider, for example, the serine proteinases. Trypsin, streptogrisin A, togavirin, IgA1-specific serine endopeptidase, hepatitis C polyprotein peptidase, etc. belong to clan SA. They all adopt a double β-barrel fold. On the other hand, those belonging to clan SC, namely, prolyl oligopeptidase, carboxypeptidase C, etc., adopt an α/β hydrolase fold (Barrett et al., 1998). The structural diversity of proteinases is equally matched by the diversity of the substrates they recognize. For a particular proteinase, only a limited extent of substrate specificity is found. Trypsin prefers to cut the peptide bond following Arg or Lys residues, while chymotrypsin cuts the peptide bonds formed by Tyr, Phe, Trp, and Leu. However, it is evident that these enzymes are capable of recognizing different substrate conformations. This indicates rugged bottoms in the binding funnels of such enzyme substrate complexes, which may, in turn, arise from the flexibility of the enzyme and of the substrate molecules. Similarly, topoisomerases I and II cleave DNA strands nonspecifically. In this case, the enzyme is also able to recognize different substrate conformations. Hence, topoisomerases are also a good example of flexible binding with the enzyme-substrate complex having rugged bottom binding funnels.

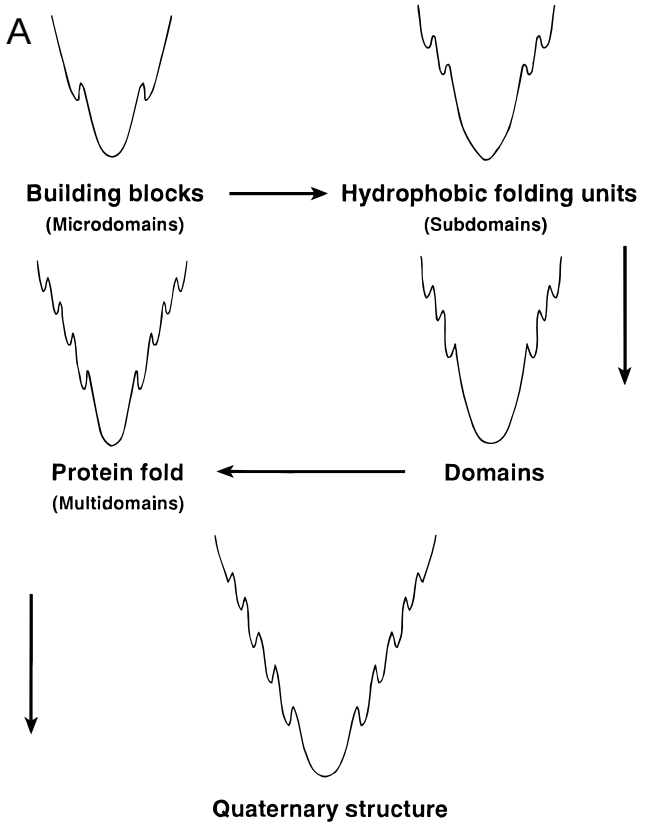

Fig. 2. The figure illustrates the hierarchical processes in folding and in binding, and the corresponding relationship between folding funnels and binding funnels, in terms of *complexity*. **A:** The funnel complexity is illustrated by the increasing number of bumps from building blocks to hydrophobic folding units (HFU), to domains, to protein fold, and finally to the quaternary structure, the outcome of a bindings process. In general, the hierarchical processes act like (consecutively) fusing two individual funnels into one. **B:** Two HFUs associate as a functional domain. **C:** Two multidomain subunits bind to form a functional heterodimer. The protein in **B** is trypsin (1SGT) with each HFU in different color. The proteins in **C** is a MHC CLASS II complex with covalently bound HB peptide (1IEA). (*Figure continues on facing page.*)

Relationship between folding funnels and binding funnels in terms of building-blocks

If the *building-block folding model*, stipulating that protein folding is the outcome of a combinatorial assembly process of a set of transient building blocks, is acceptable, folding and binding processes can be conveniently described under a common scheme (Tsai et al., 1998). In this framework, a building block in protein folding corresponds to a folded monomer in protein binding. In both cases, structural entities associate with each other, whether they are covalently connected, or unconnected. Nevertheless, while these two processes are similar, there are two inherent differences between them. First, the population times of these structural units are likely to be significantly different. A *building block* is a transient unit, which may flip between different conformations, opening, twisting, and closing (Wang & Shortle, 1996). On the other hand, a monomer is likely to be relatively stable, with a long population time. Second, intermolecular recognition between two chains in binding is a chain-linkage-free process. This is in contrast to the mutual stabilization between two building blocks, restricted by a sequential linkage. Taken together, it is not surprising to observe a similarity between folding and binding funnels. Figure 2 illustrates the similar, hierarchical processes in folding and in binding and the corresponding relationship between folding and binding funnels, in terms of complexity. Building blocks fuse into independently folding, compact, hydrophobic units. Intramolecular chain linked hydrophobic folding units bind to form structural domains, which in turn associate to form an intramolecular, multi-

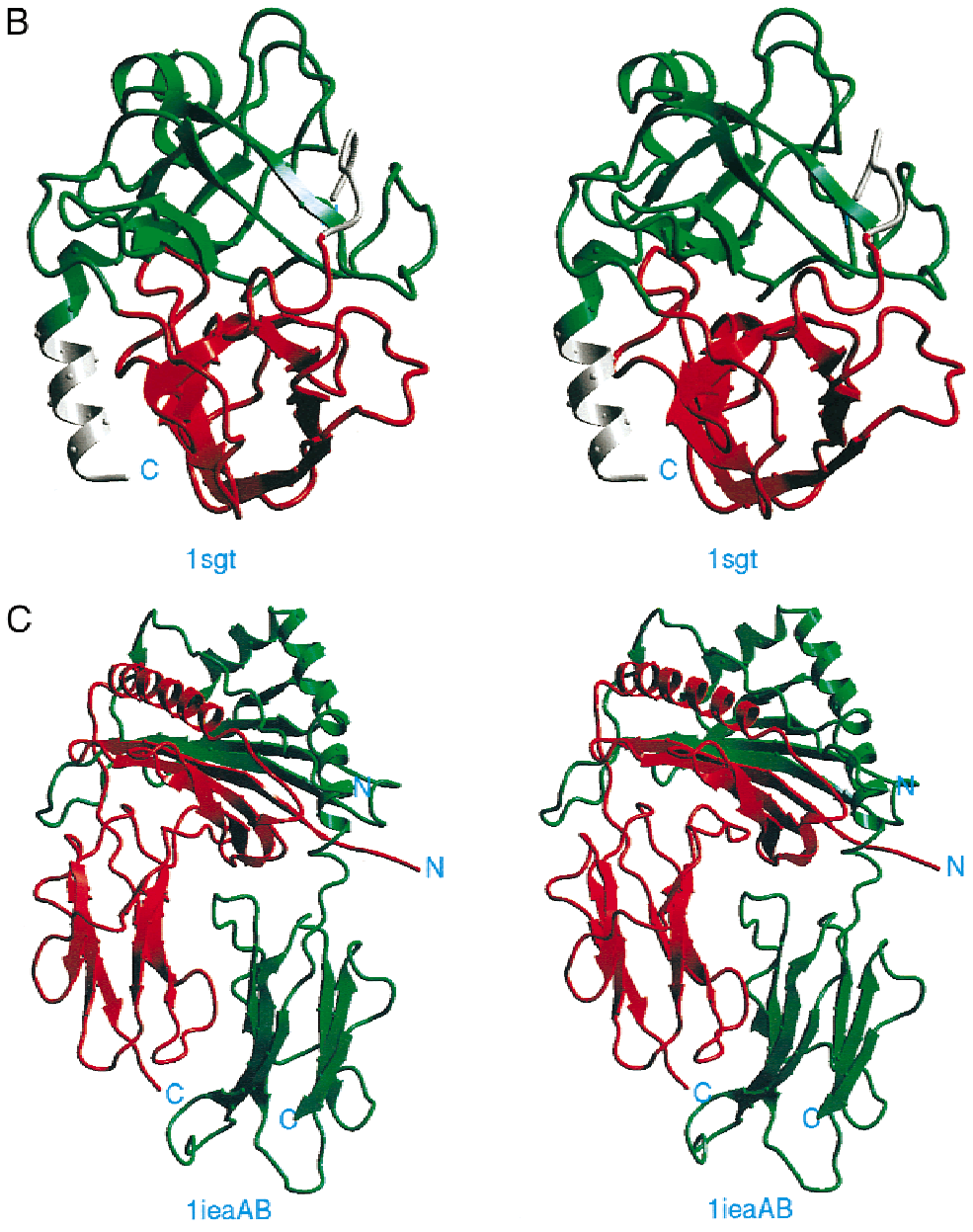

Fig. 2. *Continued.*

domain protein fold. Intermolecular binding ensues to form quaternary structure. Each of these processes entails fusing of (two) funnels into a higher dimensional funnel, regardless of whether they involve chain-connected, or unconnected, protein units. The complexity of the funnel shape increases as we climb up the hierarchical ladder.

Following the folding funnel assumption, and the building block folding model, the complexity of protein folding can be described as having two hierarchical origins. It is reasonable to assume that the formation of the building block itself already has a funnel-like shaped landscape. Consistently, based on a complete potential energy surface exploration for a four-residues peptide, Becker and Karplus (1997) have demonstrated that even such a short peptide already displays a funnel-like landscape. Hence, we may argue that crevices and bumps on the building blocks' associated funnel surfaces will be the source of folding complexity in the first step. Since a building block, by definition, is a contiguous fragment, its folding complexity depends mainly on its size and on the "foldability" of its sequence. In general, a larger fragment is likely to encounter higher folding complexity. The fact that some secondary structure formation has been observed very early in the folding process (Yue & Dill, 1995; Ballew et al., 1996; Eaton et al., 1997) indicates that the associated funnel landscape of the protein fragment is smooth with a steep descent. The non-native association between building blocks is expected to be the second and major source of folding complexity. Owing to the lower entropy of a sequentially folding protein, the surface of its folding funnel is expected to be much smoother than that of a nonsequential protein folder.

In binding, the more flexible the component monomer is, the more alike is its binding funnel to a folding funnel, especially when one of the components is disordered prior to binding. Fragment complementation studies (Tasayco & Chao, 1995; Chaffotte et al., 1997; Ladurner et al., 1997) have demonstrated that a two-fragment binding funnel resembles a folding funnel; conversely, the more nonsequential are the interactions between building blocks in protein folding, the more the folding funnel resembles a binding funnel. Folding funnels typically illustrate a single dominant valley at their bottom where the native conformation resides. On the other hand, around the bottom of two-fragment folding-binding funnels, there are valleys deep enough to be considered as traps either in folding or in binding. In protein folding, the valley represents trapped misfolded conformations; in binding, the valley corresponds to alternate binding conformations. For the particular case of sequential folding, since favorable association between building blocks reduces conformational entropy, we may expect a less bumpy surface at the bottom of the funnel. Hence, in vivo, a misfolded protein is more likely to originate from nonsequentially folding protein than from a sequential one.

Kinetics vs. thermodynamics in protein folding and binding

The thermodynamic point of view holds that the native conformations of proteins are at their global free energy minima relative to all other states having identical bonded chemistry. However, recently, it has been gradually realized that there may be large regions of conformational space that are kinetically not readily accessible. Yet, thermodynamically in these regions there might exist a more stable state. This has suggested that the folding process may be controlled both by kinetics and by thermodynamics. To a large extent, the beauty of the funnel concept lies in its being consistent with both kinetically and thermodynamically controlled folding processes. If the folding energy landscape is extremely rugged, the protein may be trapped at the conformation, which is not the global minimum. However, with time, the protein may climb out of its local minima, to reach the global minimum. A well-known example is the bovine seminal ribonuclease, existing in solution as a monomer, and eventually flipping into a domain swapped state (Piccoli et al., 1992; D'Alessio, 1995). Furthermore, it may be the case that the biologically active form of the protein is the one trapped in a local minimum. Several folding reactions appear to be determined by kinetics rather than thermodynamics (Baker & Agard, 1994). Among the four kinetically controlled folding processes reviewed by Baker and Agard, there are two such cases. The first is the heterodimerization of luciferase, which is under kinetic rather than thermodynamic control. The second is the folding of the influenza virus hemagglutinin, a trimeric viral envelope glycoprotein, which may also be under kinetic control. The funnel concept for binding processes predicts that binding may similarly be controlled both by kinetics and by thermodynamics. In both folding and binding cases, the processes initiate from higher energy and terminate in lower energy states, regardless of the pathways that are followed.

Domain swapping, misfolding, and amyloid formation

It is widely agreed that the native conformation of the protein is in general the most stable conformation. Thus, most of the population is in the native conformation. However, in the case of some proteins or their variants, like those involved in neurodegenerative diseases, a very small fraction may exist in alternate/misfolded conformations. In such proteins, once a seed of misfolded conformation forms, polymerization continues, eventually culminating in amyloid formation. As the bound-misfolded conformation is more stable than the unbound-native, the equilibrium shifts in favor of the misfolded conformer.

Viewing the folding-binding process this way explains both domain swapping and amyloid formation. The single misfolded or "open" conformers, with the domain to be swapped in a "flipped" state, are less stable than their native folded counterparts. However, in the presence of an already preformed amyloid, or domain-swapped nucleus, the flip to the conformer whose structure is complementary to the existing seed yields a more stable bound, complexed configuration. Hence, there is no conformational switch that is *induced* by the amyloid, or by an aggregate seed; rather, the less stable conformers already exist in solution, albeit in very low concentrations. Upon binding of these conformers to a pre-existing amyloid, the shifts in the equilibrium further drive the reaction, propagating the growth of the amyloid. In the case of the swapped *dimers*, these may be observed after hours of being present as nonswapped monomers (Piccoli et al., 1992; D'Alessio, 1995). Here two molecules in an "open," flipped, less stable conformation collide to form a more stable, *swapped dimer*. Similarly, swapping can also take place within the monomer, between domains, hydrophobic folding units, or other structural entities.

Folding funnels, binding funnels, and rate limiting steps

Two steps are involved in misfolding and amyloid propagation. The first is the conformational interconversion from the native to

the misfolded monomer, and the second is the binding of the misfolded conformer to the growing amyloid fiber. Can one predict the rate limiting step from the shape of the funnel? If the barriers are low, such as in cases of highly flexible proteins existing in a range of conformations, the rate limiting steps are unlikely to be the conformational interconversions. In such cases, the limiting steps might be the binding. On the other hand, if the barriers are high, the conformational interconversion may be expected to serve as the bottle neck in amyloid growth (Prusiner, 1991; Jarrett & Lansbury, 1993). However, amyloid formation is an extremely slow process. Hence, it is quite likely that the limiting step is determined by the formation of the seed, that is, the initial amyloid fiber, which in turn is determined by the conformational interconversion and binding taking place simultaneously. In the corresponding folding-binding funnel, there is an entropy barrier that cannot be overcome without the presence of amyloid. This is similar to the case of supercool water, which stays in a liquid state until a seed is dropped into it. The seed dramatically reduces the barrier height between the liquid water state and the solid ice state.

Interestingly, nucleation is also found to be the rate limiting step in protein folding. Experiment and theory are both in agreement regarding the importance of nucleation mechanisms in protein folding. The seeds for protein binding (amyloid propagation) are classical nuclei. These are well-formed elements of structure present in the ground states (final amyloid). However, the nucleation mechanisms use diffuse, extended regions, which are observed in the transition states (Fersht, 1997).

The complexity of protein folding, binding mechanisms, and the funnel shape

Given the structure of the protein monomer, we may have a notion of the shape of its funnel. The larger the complexity of its structure, the more jagged the walls of the funnel are likely to be. The structural complexity is a function of the types of interactions in the protein; namely, the complexity depends on whether the segments of the chain, which interact with each other, are sequentially or nonsequentially connected (Baldwin & Rose, 1999). In sequential folding, the shape of the funnel is the outcome of joining many microfunnels, each corresponding to a transient "building block" of the protein structure. Sequential folding has been proposed to occur more frequently in eukaryotic cells, having a slower rate of translation (Netzer & Hartl, 1997). On the other hand, in prokaryotes, where the rate of protein synthesis is considerably faster, chances of nonsequential folding are higher. Since in nonsequential folding the probability of misfolding is consequently substantially higher, there may be a more frequent need for the help of molecular chaperones. On the next hierarchical level (Tsai et al., 1998), we observe binding funnels. These similarly reflect the consequences of merging single-molecule funnels. In both folding and binding processes, the same fundamental principles are involved.

While inspection of the funnel walls narrates the story of the complexity of the folding and its potential (sequential or nonsequential) mechanism, inspection of the funnel bottoms reveals the binding mechanism. If around the bottom of the funnel of the monomer we observe a conformational ensemble, depicted as a series of minima that are close to each other on the energy landscape with low barriers separating them, we may deduce the existence of a flexible protein. On the other hand, a smooth bottom, with well-defined, deep minimum, implies a rigid molecule. Flexible molecules exhibit conformational diversity. Yet, they have often been viewed as displaying an "induced fit" binding mechanism, depending on which conformations have been crystallized in the "free," "unbound" (to its cognate ligand, though bound to its twin molecule in the crystal) form. The conformations of such molecules have also sometimes been suggested to be the outcome of "crystal effects," again, depending on which conformers have been favorable for binding in the crystal. Furthermore, folding funnels illustrating bottoms with a deep minimum, with additional minima relatively not too different in depth, and with the intervening barriers still not too high, may reflect metastable states. The corresponding proteins may be candidates for domain swapping, or for forming amyloids.

Moreover, inspection of the extent of the ruggedness around the bottom of the folding funnels enables us to have a notion of the extent of the ruggedness around the bottom of the corresponding binding funnels. In the case of a flexible monomer, a number of conformational isomers may bind a given ligand. While the stability of the resulting complexes may differ, if the single molecule conformers are relatively close on the energy landscape, the complexed conformers may similarly be conformationally and energetically quite close. On the other hand, for a rigid molecule, with a well-defined minimum (or a few minima) and a smooth bottom, a consistently similarly shaped funnel floor may be observed in the binding funnel. One may expect a considerably smaller number of complexed, favorable, bound conformations. Hence, the shape of the folding funnel of the single molecule is very instructive. The converse also holds, *if* kinetic data exist. If available, we may be able to predict the mechanism of folding, of binding, and possibly in some cases the rate limiting steps in the binding process. We shall further be able infer its binding selectivity or its nonspecificity. Thus, the concept of the funnel goes well beyond comprehension of the multiple routes going downhill, and overcoming the puzzling time scale paradox. Given the *biological considerations* of the rates of synthesis in eukaryotes and in prokaryotes, and of the function of the protein *via* its binding, the funnel concept is very useful.

Acknowledgments

We thank Dr. Jacob Maizel for encouragement and for helpful discussions. The research of R. Nussinov in Israel has been supported in part by grant number 95-00208 from BSF, Israel, by a grant from the Israel Science Foundation administered by the Israel Academy of Sciences, by the Magnet grant, by the Ministry of Science grant, and by the Tel Aviv University Basic Research and Adams Brain Center grants. This project has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under contract number NO1-CO-56000. The content of this publication does not necessarily reflect the view or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organization imply endorsement by the U.S. government.

References

Baker D, Agard DA. 1994. Kinetics versus thermodynamics in protein folding. *Biochemistry 33*:7505–7509.

Baldwin RL. 1994. Matching speed and stability. *Nature 369*:183–184.

Baldwin RL. 1995. The nature of protein folding pathways: The classical versus the new view. *J Biomol NMR 5*:103–109.

Baldwin RL, Rose GD. 1999. Is protein folding hierarchic? I. Local structure and peptide folding. *Trends Biochem Sci 24*:26–33.

Ballew RM, Sabelko J, Gruebele M. 1996. Direct observation of fast protein folding: The initial collapse of apomyoglobin. *Proc Natl Acad Sci USA 93*:5759–5764.

Barrett AJ, Rawlings ND, Woessner JF. 1998. *Handbook of proteolytic enzymes.* San Diego, California: Academic Press.
Becker OM, Karplus M. 1997. The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics. *J Chem Phys 106*:1495–1517.
Bennett MJ, Choe S, Eisenberg D. 1994. Domain swapping: Entangling alliances between proteins. *Proc Natl Acad Sci USA 91*:3127–3131.
Bennett MJ, Schlunegger MP, Eisenberg D. 1995. 3D domain swapping: A mechanism for oligomer assembly. *Protein Sci 4*:2455-2468.
Bryngelson JD, Wolynes PG. 1989. Intermediates and barrier crossing in a random energy model (with applications to protein folding). *J Phys Chem 93*:6902–6915.
Chaffotte AF, Li JH, Georgescu RE, Goldberg ME, Tasayco ML. 1997. Recognition between disordered states: Kinetics of the self-assembly of thioredoxin fragments. *Biochemistry 36*:16040–16048.
Chan HS, Dill KA. 1994. Transition states and folding dynamics of proteins and heteropolymers. *J Chem Phys 100*:9238–9257.
Chan HS, Dill KA. 1998. Protein folding in the landscape perspective: Chevron plots and non-Arrhenius kinetics. *Proteins 30*:2–33.
Cherfils J, Janin J. 1993. Protein docking algorithms: Simulating molecular recognition. *Curr Opin Struct Biol 3*:265–269.
Creighton TE. 1993. *Proteins: Structure and molecular properties, 2nd ed.* New York: WH Freeman and Company.
D'Alessio G. 1995. Oligomer evolution in action. *Nat Struct Biol 2*:11–13.
Dill KA. 1999. Polymer principles and protein folding. *Protein Sci 8*:1166–1180.
Dill KA, Chan HS. 1997. From Levinthal to pathways to funnels. *Nat Struct Biol 4*:10–19.
Dill KA, Phillips, AT, Rosen JB. 1997. Protein structure and energy landscape dependence on sequence using a continuous energy function. *J Comp Biol 4*:227–239.
Dunbrack RL, Gerloff DL, Bower M, Chen X, Lichtarge O, Cohen FE. 1997. Meeting review: The second meeting on the critical assessment of techniques for protein structure prediction (CASP2), Asilomar, California, December 13–16, 1996. *Folding Design 2*:R27–R42.
Eaton WA, Munoz V, Thompson PA, Chan CK, Hofrichter J. 1997. Submillisecond kinetics of protein folding. *Curr Opin Struct Biol 7*:10–14.
Fersht AR. 1997. Nucleation mechanisms in protein folding. *Curr Opin Struct Biol 7*:3–9.
Fischer D, Lin SL, Wolfson HJ, Nussinov, R. 1995. A geometry-based suite of molecular docking processes. *J Mol Biol 248*:459–477.
Fischer E. 1894. *Ber Dt Chem Ges 27*:2985–2991.
Foote J, Milstein C. 1994. Conformational isomerism and the diversity of antibodies. *Proc Natl Acad Sci USA 91*:10370–10374.
Frauenfelder H, Leeson DT. 1998. The energy landscape in non-biological molecules. *Nat Struct Biol 5*:757–759.
Frauenfelder H, Sligar SG, Wolynes PG. 1991. The energy landscapes and motions of proteins. *Science 254*:1598–1603.
Gruebele M, Wolynes P. 1998. Satisfying turns in folding transitions. *Nat Struct Biol 5*:662–665.
Jarrett JT, Lansbury PT Jr. 1993. Seeding "one-dimensional crystallization" of amyloid: A pathogenic mechanism in Alzheimer's disease and scrapie? *Cell 73*:1055–1058.
Karplus M. 1997. The Levinthal paradox: yesterday and today. *Folding Design 2*:S69–S75.
Karplus M, Sali A, Shakhnovitch E. 1995. Comment: Kinetics of protein folding. *Nature 373*:664–665.
Karplus M, Shakhnovitch E. 1992. Protein folding: Theoretical studies of thermodynamics and dynamics. In: Creighton T, ed. *Protein folding*. New York: WH Freeman & Sons. pp 127–195.
Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem AA, Aflalo C, Vakser IA. 1992. Molecular surface recognition: Determination of geometric fit between protein and their ligands by correlation techniques. *Proc Natl Acad Sci USA 89*:2195–2199.
Koshland DE Jr. 1958. Application of a theory of enzyme specificity to protein synthesis. *Proc Natl Acad Sci USA 44*:98–123.
Ladurner AG, Itzhaki LS, Gay FDP, Fersht AR. 1997. Complementation of peptide fragments of the single domain protein chymotrypsin inhibitor 2. *J Mol Biol 273*:317–329.
Lazaridis T, Karplus M. 1997. "New view" of protein folding reconciled with the old through multiple unfolding simulations. *Science 278*:1928–1931.
Lee AY, Gulnik SV, Erickson JW. 1998. Conformational switching in an aspartic proteinase. *Nat Struct Biol 5*:866–871.
Levinthal C. 1969. How to fold graciously. In: Debrunner P, Tsibris JCM, Munck E, eds. *Mossbauer spectroscopy in biological systems* (Proceedings of a meeting held at Allerton House, Monticello, Illinois). Urbana, Illinois: University of Illinois Press. p 22.
Martinez JC, Pisabarro MT, Serrano L. 1998. Obligatory steps in protein folding and the conformational diversity of the transition state. *Nat Struct Biol 5*:721–729.
Miller DW, Dill KA. 1997. Ligand binding to proteins: The binding landscape model. *Protein Sci 6*:2166–2179.
Muller Y, Kelley RF, De Vos AM. 1998. Hinge bending within the cytokine receptor superfamily revealed by the 2.4 Å crystal structure of the extracellular domain of the rabbit tissue factor. *Protein Sci 7*:1106–1115.
Netzer WJ, Hartl FU. 1997. Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature 388*:343–349.
Norel R, Lin SL, Wolfson H, Nussinov R. 1994. Shape complementarity at protein–protein interfaces. *Biopolymers 34*:933–940.
Norel R, Lin SL, Wolfson H, Nussinov R. 1995. Molecular surface complementarity at protein–protein interfaces: The critical role played by surface normals at well placed, sparse, points in docking. *J Mol Biol 252*:263–273.
Northrup SH, Erickson HP. 1992. Kinetics of protein–protein association explained by Brownian dynamics computer simulation. *Proc Natl Acad Sci USA 89*:3338–3342.
Onuchic JN, Wolynes PG, Luthey-Schulten Z, Socci ND. 1995. Towards an outline of the topography of a realistic protein folding funnel. *Proc Natl Acad Sci USA 92*:3626–3630.
Perona JJ, Craik CS. 1995. Structural basis of substrate specificity in the serine proteases. *Protein Sci 4*:337–360.
Piccoli R, Tamburrini M, Piccialli G, Di Donato A, Parente A, D'Alessio G. 1992. The dual-mode quaternary structure of seminal rnase. *Proc Natl Acad Sci USA 89*:1870–1874.
Prusiner SB. 1991. Molecular biology of prion diseases. *Science 252*:1515–1522.
Tasayco ML, Chao K. 1995. NMR study of the reconstitution of the β-sheet of thioredoxin by fragment complementation. *Proteins SFG 22*:41–44.
Tsai CJ, Lin SL, Wolfson, H, Nussinov R. 1996. Protein–protein interfaces: Architectures and interactions in protein–protein interfaces and in protein cores. Their similarities and differences. *Critic Rev Biochem Mol Biol 31*:127–152.
Tsai CJ, Lin SL, Wolfson, H, Nussinov R. 1997a. Studies of protein–protein interfaces: Statistical analysis of the hydrophobic effect. *Protein Sci 6*:53-64.
Tsai CJ, Maizel JV Jr, Nussinov R. 1999. Distinguishing between sequential and non-sequentially folded protein: Implications for folding and misfolding. *Protein Sci*. In press.
Tsai CJ, Nussinov R. 1997. Hydrophobic folding units derived from dissimilar monomer structures and their interactions. *Protein Sci 6*:24–42.
Tsai CJ, Xu D, Nussinov R. 1997b. Structural motifs at protein–protein interfaces: Protein cores versus two-state and three-state model complexes. *Protein Sci 6*:1793–1805.
Tsai CJ, Xu D, Nussinov R. 1998. Protein folding via binding, and vice versa. *Folding Design 3*:R71–R80.
Verkhivker GM, Rejto PA, Gehlhaar DK, Freer ST. 1996. Exploring the energy landscapes of molecular recognition by a genetic algorithm: Analysis of the requirements for robust docking of HIV-1 protease and FKBP-12 complexes. *Proteins SFG 250*:342–352.
Wales DJ, Miller MA, Walsh TR. 1998. Archetypal energy landscapes. *Nature 394*:758–760.
Wang Y, Shortle D. 1996. A dynamic bundle of four adjacent hydrophobic segments in the denatured state of staphylococcal nuclease. *Protein Sci 5*:1898–1906.
Wedemayer GJ, Patten PA, Wang LH, Schultz PG, Stevens RC. 1997. Structural insights into the evolution of an antibody combining site. *Nat Struct Biol 276*:1665–1669.
Wells JA. 1996. Binding in the growth hormone receptor complex. *Proc Natl Acad Sci USA 93*:1–6.
Wolynes PG, Onuchic JN, Thirumalai D. 1995. Navigating the folding routes. *Science 267*:1619–1620.
Yue K, Dill KA. 1995. Forces of tertiary structural organization in globular proteins. *Proc Natl Acad Sci USA 92*:146–150.
Xu D, Lin SL, Nussinov R. 1997. Protein binding versus protein folding: The role of hydrophilic bridges in protein association. *J Mol Biol 265*:68–84.
Xu D, Tsai CJ, Nussinov R. 1998. Mechanism and evolution of protein dimerization. *Protein Sci 7*:533–544.
Zhang C, Chen J, DeLisi C. 1999. Protein–protein recognition: Exploring the energy funnels near the binding sites. *Proteins SFG 34*:255–267.